Toward an individual-specific model of impaired speech intelligibility

Van Summers[1], Matthew J. Makashay[1], Elena Grassi[1], Ken W. Grant[1], Joshua G. W. Bernstein[1], Brian E. Walden[1], Marjorie R. Leek[2], and Michelle R. Molis[2]

[1] *Walter Reed Army Medical Center, Army Audiology and Speech Center, Washington, DC, USA*

[2] *VA RR&D National Center for Rehabilitative Auditory Research, Portland, OR, USA*

Hearing-impaired listeners with similar quiet thresholds often show very different real-world speech intelligibility deficits in listening situations involving competing sounds. The current study is part of larger research project focused on: (1) examining how, and to what extent, these between-subject differences in speech recognition relate to differences in suprathreshold auditory functioning, and (2) on generating accurate, individualized models to predict auditory and auditory-visual speech recognition by hearing-impaired listeners in adverse listening conditions. Individual hearing-impaired and normal-hearing listeners are being tested on a range of psychoacoustic tasks intended to characterize auditory processing sensitivity along a variety of dimensions (frequency selectivity, peripheral compression, traveling wave dispersion, inner hair cell status, spectral and temporal modulation sensitivity and fine-structure processing). Here we report estimates of frequency selectivity and peripheral compression for hearing impaired listeners with similar audiograms and very different performance on speech in noise. Comparisons between hearing-impaired and normally-hearing subjects on psychoacoustic measures and their relation to speech recognition in noise will be discussed.

INTRODUCTION

Hearing-impaired (HI) listeners with similar audiograms may show large differences in real-world speech recognition performance in noisy or reverberant listening conditions. For listeners with mild to moderate hearing loss, speech recognition performance in noise may be more closely related to suprathreshold processing than audibility (Plomp, 1978; Glasberg and Moore, 1989). However, some previous studies investigating the relationship between frequency selectivity and speech perception in HI listeners have reported fairly low correlations once audibility is partialled out (e.g. Glasberg and Moore, 1989). Possible reasons for this may include (1) the conflation of audibility and frequency selectivity, and (2) the focus on filter bandwidth alone and not the reduction in cochlear nonlinearity that accompanies reduced bandwidth. To address the audibility issue, speech is presented at a high level (92 dB SPL) in a background noise such that the signal-to-noise ratio (SNR) and not the absolute level will tend to drive performance. As HI listeners show particular deficits in benefiting from masker modulation (Festen and Plomp, 1990), speech intelligibility was also meas-

Auditory signal processing in hearing-impaired listeners. 1st International Symposium on Auditory and Audiological Research (ISAAR 2007). T. Dau, J. M. Buchholz, J. M. Harte, T. U. Christiansen (Eds.). ISBN: 87-990013-1-4. Print: Centertryk A/S.

ured in a speech-modulated noise condition. Finally, to investigate the possible role of the cochlear nonlinearity in speech perception, a forward-masking paradigm with variable masker-delay is employed (Nelson, *et al.*, 2001; Lopez-Poveda *et al.*, 2003), in addition to frequency selectivity measures using a notched-noise paradigm (Rosen and Baker, 1994). The results will be used, in combination with other psychoacoustic measures to be performed in the future characterizing other aspects of peripheral and central processing, to develop individual-specific models of auditory processing underlying speech recognition performance.

METHODS

Participants

Data are reported for three hearing-impaired (HI) and four normal hearing (NH) listeners. Audiometric thresholds for all listeners appear in Figure 1. HI listeners (closed symbols) had similar absolute thresholds above 500 Hz. At 250 and 500 Hz, HI7 had thresholds in the normal range (≤ 10 dB HL) while the other HI listeners had slightly elevated thresholds. Data for the three HI listeners and one or more NH listeners are reported for each experimental task (below).

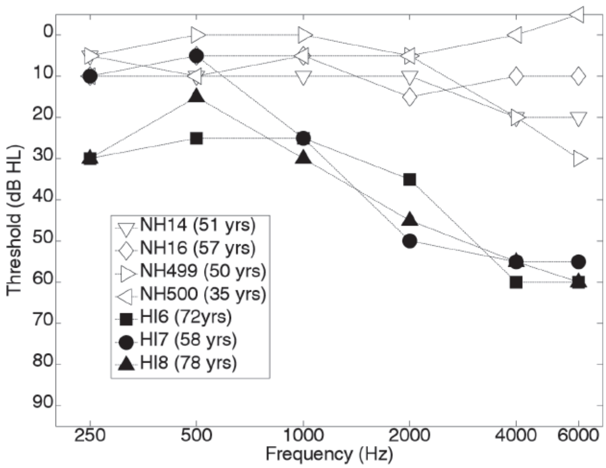

Fig. 1: Audiometric thresholds for HI and NH listeners.

Speech intelligibility

Stimuli

Sentences from the Institute of Electrical and Electronic Engineers sentence corpus (IEEE, 1969), spoken by a female talker, were used as target stimuli. Modulated (MOD) and steady-state (SS) speech-shaped noise were used to mask target sentences. Masker stimuli had long-term spectra matching the average spectrum of the IEEE sentences. For the MOD condition, the SS noise was modulated based on the envelope of a two-talker babble (Festen and Plomp, 1990). For the MOD testing, in addition to the modulated noise masker, a low-level SS noise was presented (-20 dB SNR) to limit the available speech cues to just the upper portion of the speech dynamic range. This was done to better equate the available range of speech cues between NH and HI listeners where the low-level portions of the speech might have been audible for NH listeners but not for HI listeners.

Procedure

Target sentences and maskers were lowpass filtered at 8 kHz and digitally mixed at the appropriate SNR before being converted to analog form (TDT RP2.1), and led via a headphone buffer (TDT HB7) to an earphone (Sennheiser HD580). Subjects were seated in a sound-treated booth. Target and masker stimuli were presented monaurally. The ear tested was the same ear tested in the psychoacoustic tasks described below. Subjects were instructed to attend to each spoken sentence and repeat it back as accurately as possible. Subjects' verbal responses were scored for the number keywords correct (out of five per sentence). Each subject was given 30 minutes of practice with feedback provided in the form of target sentence orthographically displayed on a monitor. No feedback was provided during the testing phase. Target sentences were presented at 92 dB SPL. The noise level varied across sentence lists to produce four SNR conditions: -6, -3, 0, and +3 dB. Three target lists (30 sentences, 150 key-words) were assigned to each noise type and SNR condition at the start of the experiment so that the same target lists were used to test each listening condition for all listeners.

Psychophysical tasks

Tone detection in notched-noise

Stimuli

Probe stimuli were 400-ms sinusoids with frequencies (f_p) of 0.5, 1, 2 or 4 kHz. Probe levels were fixed within a block of trials, and varied across experimental blocks (see Fig. 3 for specific probe levels for each subject). Notched-noise stimuli (composed of a pair of bandpass noises separated by spectral gap or "notch") were used as maskers. Edge frequencies of the spectral notch in the masker stimuli are characterized in terms of normalized units relative to the f_p. For example, a .2.4 notched noise would contain a spectral notch between $-0.2f_p$ and $+0.4f_p$. Six notch conditions, identified by Stone *et al.* (1992) as sufficient to characterize sharpness and asymmetry of peripheral filtering, were tested (.0.0, .1.1, .2.2, .4.4, .2.4, .4.2). Outer edges of the notched-noise maskers were fixed at $\pm 0.8f_p$. Notched-noise maskers were 500 ms in duration and probes were temporally centered in the maskers.

Procedure

Masker levels were varied adaptively in a two-alternative forced-choice procedure to determine the masker level supporting 79% correct probe detection (Levitt, 1971). For each combination of probe level, probe frequency, and notched-noise masker, at least two threshold estimates were obtained. If two threshold estimates for the same condition differed by 4 dB or more, a third estimate was collected and the mean of all three estimates was used.

Analysis

The PolyFit procedure (Rosen and Baker, 1994) was used to characterize auditory filter shapes at different probe levels. Fitting was based on a version of the rounded exponential (roex) function describe by Patterson *et al.* (1982), with three parameters: p (the slopes of the filter tip), w (the breakpoint between the filter tip and the low-frequency tail, and t (the slope of the low-frequency tail). The w and t parameters were only applied to the low-frequency side of the filter, with the upper side described only by the constant slope p. The low-frequency filter parameters were allowed to vary as a linear function of probe level (Rosen *et al.*, 1998).

Temporal masking curves (TMCs)

Stimuli

Probe stimuli were 20-ms sinusoids including 4-ms onset and offset ramps, with f_p's of 0.5, 1, 2, and 4 kHz. Probe level was fixed at 10 dB SL. Masker stimuli were 108-ms sinusoids including 4-ms ramps, with frequencies (f_m's) of 1.0 f_p or .55 f_p. Maskers and probes were presented in a forward masking paradigm with the delay (T) between masker offset and probe onset ranging from 10 to 100 ms.

Procedure

On a given block of trials, f_p, f_m and T were held constant, and the masker level varied adaptively in a two-alterntive forced –choice procedure to determine the masker level supporting 79% correct probe detection (Levitt, 1971). Two threshold estimated were obtained, and averaged, for each combination of f_p, f_m and T.

Analysis

The TMC task allows estimates of cochlear compression under the assumption that internal excitation from a forward masker decays exponentially with time, at a rate that is independent of frequency and level. As masker-probe delay increase, the decay creates the need for a proportionally higher masker level to mask the same probe signal. With compressive internal processing, an additional increase in masker level will be necessary to produce the same level of internal excitation. In the current analysis, for each listener, "off-frequency" masking results for the 4kHz probe (f_m =2200 Hz) were used as "reference", to estimate temporal decay of forward masking and deviations from this reference were interpreted as cochlear compression.

RESULTS

Speech Recognition

The three HI listeners showed clear differences in performance on the speech tasks (Figure 2). HI6 and HI8 had similar speech scores to one another and lower scores than HI7 for all test conditions. Speech scores for HI7 were similar to scores for listener NH14 for the steady-state noise masker. NH14 outperformed HI7 in modulated maskers. In general, the HI listeners showed little or no benefit from modulation in competing maskers, consistent with previous findings (e.g. Festen and Plomp, 1990).

Speech Intelligibility Index scores (ANSI, 1997) for the sentence stimuli were nearly identical for the three HI listeners, suggesting audibility differences had little effect on speech scores. This suggests that the large difference in speech performance between HI7 and the other HI listeners indicate differences in peripheral and/or central auditory processing that are not reflected in the audiogram. The notched-noise and TMC results, reported next, allow an examination of whether performance on these psychoacoustic tasks (frequently linked to outer hair cell status) are associated with performance on the speech tasks.

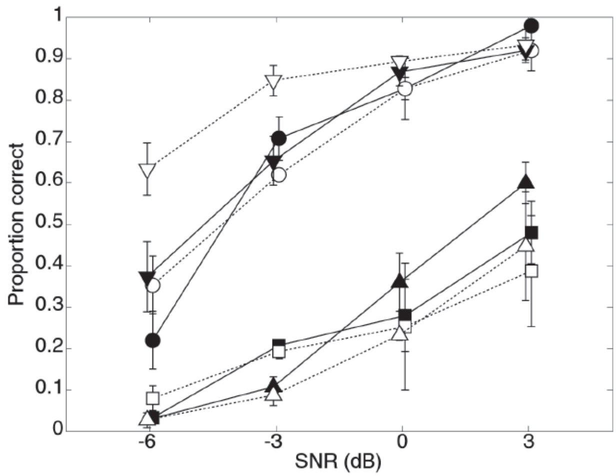

Fig. 2: Speech scores in steady-state noise (solid symbols) and in modulated noise (open symbols) for HI listeners and NH14. Symbols indicate same listeners as in Figure 1.

Tone detection in notched-noise

Equivalent rectangular bandwidths (ERBs) of auditory filter estimates for the HI subjects and two NH subjects appear in Figure 3. ERBs at 500 Hz are similar for HI and NH listeners, consistent with the near-normal hearing thresholds for the HI subjects at 500 Hz. In addition, all subjects showed fairly similar increases in ERB with probe level, suggesting similar reductions in frequency selectivity with level at 500 Hz. At 1000 Hz, where audiometric thresholds for the HI listeners were slightly higher than normal, ERBs were slightly broader than for NH listeners. All subjects again showed increases in ERB with level. ERBs at 2000 and 4000 Hz were clearly greater for the three HI listeners, consistent with the re elevated pure tone thresholds at these frequencies.

At 4000 Hz, HI listeners had similar pure tone thresholds but different ERBs. Specifically, HI6 showed extremely broad filters (essentially no frequency selectivity). The much broader ERBs for HI6 than HI7 may have contributed to the large differences in speech recognition for these listeners (Fig. 2). However, HI7 also had much higher speech scores than HI8 but had only slightly narrower ERBs at 4 kHz. Frequency selectivity at 4kHz was similar across presentation levels HI listeners suggesting linear cochlear processing.

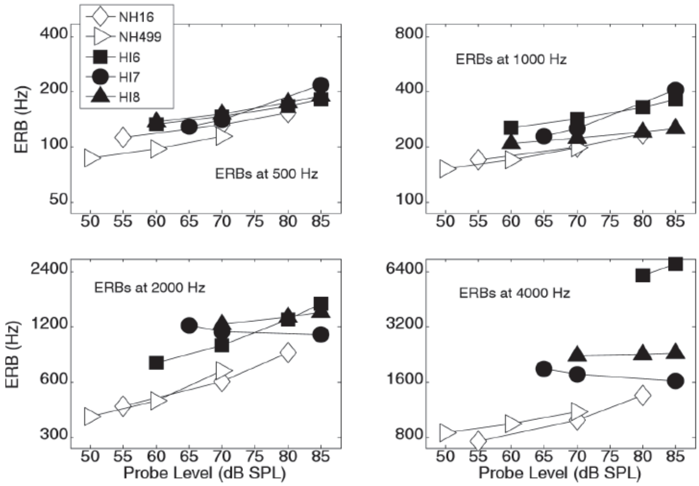

Fig 3: Equivalent Rectangular Bandwidths (ERBs) of auditory filters at four frequencies for HI subjects and two NH subjects.

TMCs

Representative TMCs from one NH and one HI subject (NH 500 and HI6) for on- and off-frequency maskers of a 4000-Hz probe appear in the left-hand panel of Figure 4. For each subject, the off-frequency masking results were treated as a linear reference allowing estimates of compression in on-frequency cochlear response growth (Nelson *et al.*, 2001). For NH500, slopes of the on-frequency TMCs were less steep (and more similar to the slopes of the off-frequency TMCs) when on-frequency masker levels were below 40 dB SPL and above 90 dB SPL than at intermediate levels. This suggests more linear processing at very low and very high input levels than at moderate levels, consistent with previous results (e.g. Lopez-Poveda *et al.*, 2003). For HI6, slopes of on- and off-frequency TMC curves were similar across masker delays, suggesting linear cochlear processing at 4000 Hz. For this listener, the on-frequency masker was consistently less effective than the off-frequency masker which may indicate a cochlear dead region for the BM region tuned to 4000 Hz. However, Psychophysical Tuning Curve results (not shown) did not support the presence of a dead region at 4000 Hz for this listener.

Estimates of peripheral compression at four frequencies (based on the TMC data) are shown for the HI subjects and NH500 in the right-hand panel of Figure 4. For NH listeners, previous studies report estimates of on-frequency response growth between .20 and .33 dB/dB at moderate levels (Lopez-Poveda *et al.*, 2003). The data for NH500 are generally within this range. Compression estimates for the HI listeners also fall within this range at 500 and 1000 Hz; and at 2000 Hz for HI6 and HI7. HI8 demonstrated more linear response growth at 2000 Hz. All three HI listeners showed similar and fairly linear growth estimates at 4000 Hz. The TMC results do not appear to distinguish HI7 (who outperformed the other HI listeners on the speech tasks) from the other HI listeners.

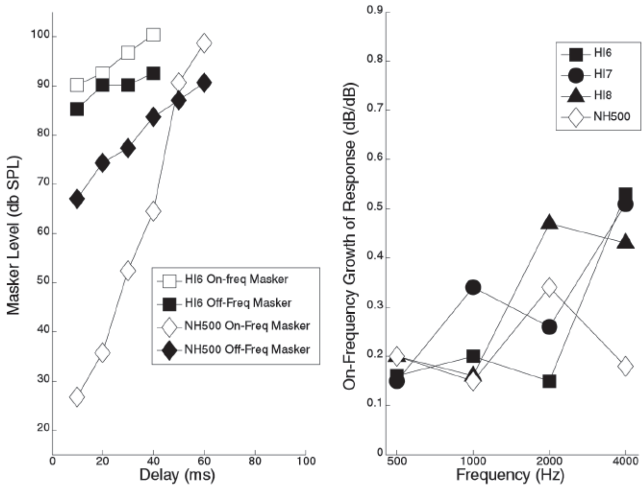

Fig. 4: TMCs for a 4000-Hz probe: subjects HI6 and NH500 (left-hand panel). Derived on-frequency response growth estimates for HI listeners and NH500 at four probe frequencies (right hand panel).

DISCUSSION

The current psychoacoustic measures indicate only small differences between the HI listeners in frequency selectivity and compression. Subject HI7, who outperformed the other HI listeners in the speech tasks, had slightly lower absolute thresholds at 250 and 500 Hz and greater frequency selectivity at 4000 Hz. The psychoacoustic data indicate more impaired processing at 4000 Hz for subject HI6 than the other listeners. However, HI6 and HI8 performed similarly on speech in noise. Efforts are underway to model the internal representations of speech in noise for these listeners. Modeling combines a peripheral model similar to the Dual-Resonance Nonlinear model described by Lopez-Poveda and Meddis (2001) with a cortical model developed by Shamma and colleagues (Chi *et al.*, 1999; Elhilali *et al.*, 2003). The cortical modeling highlights spectral and temporal modulations in the signal and compares the outputs of a “clean” speech signal through normal hearing earlier stages with the speech signal hypothesized to emerge from earlier stages which model impaired processing by individual listeners. The approach is similar to recent modeling efforts reported by Zilany and Bruce (2007) but is focused on modeling auditory processing in individual subjects rather characterizing group performance.

The current psychophysical testing did not include tasks that directly examined temporal processing. It is possible that temporal processing differences across the three HI subjects contributed to the differences in speech recognition performance. However, the inability of any of the three HI listeners to benefit from temporal modulation in the masker stimuli in the speech testing may indicate impaired temporal processing for all three listeners. Speech perception may be affected by diminished temporal fine-structure processing abilities (Buss *et al.*, 2004; Lorenzi *et al.*, 2006). We plan to incorporate a frequency-modulation detection task (Moore and Szrodska, 2002) in the test battery to address this possibility.

Changes in central, cognitive processing with age may contribute to poor speech recognition in older listeners (Gordon-Salant & Fitzgibbons, 1997). HI7 was younger than the other HI listeners, and outperformed them in speech testing. This may indicate that central processing differences contributed to the speech results. Along with the peripheral measures used to individualize the auditory models of these listeners, we will be testing sensitivity to spectral and temporal modulation using "ripple stimuli" (Chi *et al.*, 1999). These measures should provide an indication of whether more central processing differences involving modulation sensitivity contribute to differences in speech performance.

CONCLUSIONS

Preliminary results indicate that measures of peripheral frequency selectivity and cochlear compression may not account for individual differences in speech intelligibility in stationary and speech-modulated noise for HI listeners. Therefore, in addition to these measures, estimates of fine structure processing, cochlear dead regions, and broadband temporal and spectral modulation processing will be incorporated into individual-specific models of auditory function in an effort to account for observed differences in speech intelligibility performance.

ACKNOWLEDGMENTS

This research was partially supported by Cooperative Research and Development Agreements between the Clinical Investigation Regulatory Office, U.S. Army Medical Department and the Oticon Foundation, Copenhagen, Denmark. The authors would like to thank the Department of Clinical Investigation at the Walter Reed Army Medical Center for their administrative support and Institutional Review of research involving human subjects under work unit #06-25025. The opinions and assertions presented are the private views of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Army, or the Department of Defense.

REFERENCES

ANSI (**1997**). American Standard Methods for Calculation of the Speech Intelligibility Index, ANSI S3.5-1997, (American National Standards Institute, New York).

Buss, E., Hall, J. W. and Grose, J. H. (**2004**). "Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss," Ear. Hear. **25**, 242-250.

Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation tranfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719-2732.

Elhilali, M., Chi, T., and Shamma, S. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Comm. **41**, 331-348.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725-1736.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1-25.

Gordon-Salant, S., and Fitzgibbons, P., (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Sp. Lang. Hear. Res. **40**, 423-431.

Institute of Electrical and Electronic Engineers. (**1969**). "IEEE Recommended Practice for Speech Quality Measures. IEEE, New York."

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467-477.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). "A human nonlinear cochlear filter bank," J. Acoust. Soc. Am. **110**, 3107-3318.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). "Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing," J. Acoust. Soc. Am. **113**, 951-960.

Lorenzi, C., Gilbert, G., Carn, H. Garnier, S. and Moore, B. C. J. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Nat. Acad. Sci. **103**, 18666-18669.

Moore, B. C. J. and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation.," J. Acoust. Soc. Am. **111**, 327-335.

Nelson, D. A., Schroder, A. C., and Wojtczak, M. (**2001**). "A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 2045-2064.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788-1803.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533-549.

Rosen, S., and Baker, R.J. (**1994**). "Characterizing auditory filter nonlinearity," Hear. Res. **73**, 231-243.

Rosen, S., Baker, R.J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539-2550.

Stone, M. A., Glasberg, B. R., and Moore, B. C. J. (**1992**). "Simplified measurement of auditory filter shapes using the notched-noise method," Br. J. Audiol. **26**, 329-334.

Zilany, M. S. A., and Bruce, I. C. (**2007**). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," Proceedings of the 3rd International IEEE/EMBS Conference on Neural Engineering, Kohala Coast, HI, 481-485.